

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
8 November 2007 (08.11.2007)

PCT

(10) International Publication Number
WO 2007/127501 A2

(51) International Patent Classification:
C12N 15/29 (2006.01)

(21) International Application Number:
PCT/US2007/060235

(22) International Filing Date: 8 January 2007 (08.01.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/757,544 9 January 2006 (09.01.2006) US

(71) Applicant (*for all designated States except US*): CERES,
INC. [US/US]; 1535 Rancho Conejo Boulevard, Thousand
Oaks, California 91320-1440 (US).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): VANDINH DANG,
David [US/US]; 274 Saint Croix Court, Oak Park, Cali-
fornia 91377 (US).

(74) Agent: LUNDQUIST, Ronald C.; Fish & Richardson
P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-1022
(US).

(81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,

AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS,
JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS,
LT, LU, LV, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO,
RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every
kind of regional protection available*): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT,
RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA,
GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a
patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii))*

Published:

- *with declaration under Article 17(2)(a); without abstract;
title not checked by the International Searching Authority*

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: NOVEL ENDOSPERM REGULATORY REGIONS

(57) Abstract:



WO 2007/127501 A2

NOVEL ENDOSPERM REGULATORY REGIONS

5 CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority under 35 U.S.C. § 119 to U.S. Provisional Application Ser. No. 60/757,544, filed January 9, 2006, the entire contents of which are incorporated herein in their entirety.

10 TECHNICAL FIELD

This document provides compositions and methods involved in regulating gene expression in eukaryotic organisms (*e.g.*, plants).

15 INCORPORATION-BY-REFERENCE & TEXT

The material on the accompanying compact discs is hereby incorporated by reference into this application. The accompanying compact discs all contain one identical file, 18207-011WO1-Sequence.txt, which was created on January 8, 2007. The file named 18207-011WO1-Sequence.txt is 8 KB. The file can be
20 accessed using Microsoft Word on a computer that uses Windows OS.

BACKGROUND

An essential element for genetic engineering of plants is the ability to express genes using various regulatory regions. The expression pattern of a
25 transgene, conferred by a regulatory region is critical for the timing, location, and conditions under which a transgene is expressed, as well as the intensity with which the transgene is expressed in a transgenic plant. There is continuing need for suitable regulatory regions that can facilitate transcription of sequences that are operably linked to the regulatory region.

30

SUMMARY

This document provides compositions and methods involving regulatory regions having the ability to direct transcription in eukaryotic organisms (*e.g.*, plants). For example, this document provides regulatory regions having the

ability to direct transcription in plant ovules prior to fertilization and in seeds

during early stages of development. Also provided herein are nucleic acid constructs, plant cells, and plants containing such regulatory regions; methods of producing plant cells and plants containing such regulatory regions; and methods of using such regulatory regions to express polynucleotides in plants and to alter the phenotype of plant cells. Regulatory regions that direct transcription during seed development can be used, for example, to manipulate genomic imprinting in crop plants, resulting in enhanced seed development and increased yield.

In one embodiment, an isolated nucleic acid including a regulatory region having 80 percent or greater sequence identity to SEQ ID NO:1 or SEQ ID NO:2 is provided. The regulatory region directs transcription, in a plant ovule within 24 hours post-fertilization, of an operably linked heterologous polynucleotide. The sequence identity can be 85 percent or greater or 90 percent or greater. The regulatory region can include an intron.

In another embodiment, an isolated nucleic acid including a regulatory region having 80 percent or greater sequence identity to SEQ ID NO:1 or SEQ ID NO:2 is provided. The regulatory region directs transcription, in an unfertilized plant ovule, of an operably linked heterologous polynucleotide. The sequence identity can be 85 percent or greater or 90 percent or greater. The regulatory region can include an intron.

In another embodiment, a nucleic acid construct is provided. An isolated nucleic acid including a regulatory region described above that is operably linked to a heterologous polynucleotide is included in the nucleic acid construct. The heterologous polynucleotide can include a nucleic acid sequence encoding a polypeptide. The heterologous polynucleotide can be in an antisense orientation relative to the regulatory region. The heterologous polynucleotide can be transcribed into an antisense RNA capable of inhibiting expression of a DNA methyltransferase. The heterologous polynucleotide can be transcribed into an interfering RNA. The heterologous polynucleotide can be transcribed into an interfering RNA against a DNA methyltransferase.

In another embodiment, a transgenic plant or plant cell is provided. The plant or plant cell is transformed with an isolated nucleic acid including a regulatory region operably linked to a heterologous polynucleotide. The heterologous polynucleotide can include a nucleic acid sequence encoding a

polypeptide. The heterologous polynucleotide can be in an antisense orientation relative to the regulatory region. The heterologous polynucleotide can be transcribed into an interfering RNA.

In another embodiment, a method of producing a transgenic plant is provided. The method can include (a) introducing into a plant cell an isolated polynucleotide including an isolated nucleic acid including a regulatory region operably linked to a heterologous polynucleotide, and (b) growing a plant from the plant cell. The heterologous polynucleotide can include a nucleic acid sequence encoding a polypeptide. The heterologous polynucleotide can be in an antisense orientation relative to the regulatory region. The heterologous polynucleotide can be transcribed into an interfering RNA.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Although methods and materials similar or equivalent to those described herein can be used to practice the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

The details of one or more embodiments of the invention are set forth in the accompanying description below. Other features, objects, and advantages of the invention will be apparent from the description and from the claims.

DETAILED DESCRIPTION

The invention features isolated nucleic acids comprising regulatory regions. The terms “nucleic acid” and “polynucleotide” are used interchangeably herein, and refer to both RNA and DNA, including cDNA, genomic DNA, synthetic DNA, and DNA (or RNA) containing nucleic acid analogs. Polynucleotides can have any three-dimensional structure. A nucleic acid can be double-stranded or single-stranded (*i.e.*, a sense strand or an antisense strand). Non-limiting examples of polynucleotides include genes, gene fragments, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, siRNA, micro-RNA, ribozymes, cDNA, recombinant polynucleotides,

branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers, as well as nucleic acid analogs.

An isolated nucleic acid can be, for example, a naturally-occurring DNA molecule, provided one of the nucleic acid sequences normally found immediately flanking that DNA molecule in a naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a DNA molecule that exists as a separate molecule (*e.g.*, a chemically synthesized nucleic acid, or a cDNA or genomic DNA fragment produced by the polymerase chain reaction (PCR) or restriction endonuclease treatment) independent of other sequences. An isolated nucleic acid also refers to a DNA molecule that is incorporated into a vector, an autonomously replicating plasmid, a virus, a bacterium, or into the genome of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include an engineered nucleic acid such as a DNA molecule that is part of a hybrid or fusion nucleic acid. A nucleic acid existing among hundreds to millions of other nucleic acids within, for example, cDNA libraries or genomic libraries, or gel slices containing a genomic DNA restriction digest, is not to be considered an isolated nucleic acid.

Regulatory regions

A regulatory region described herein is a nucleic acid that can direct transcription when the regulatory region is operably linked 5' to a heterologous nucleic acid. As used herein, "heterologous nucleic acid" refers to a nucleic acid other than the fertilization-independent endosperm 2 (FIE2) polypeptide coding sequence from rice. The term "operably linked" refers to positioning of a regulatory region and a transcribable sequence in a nucleic acid so as to allow or facilitate transcription of the transcribable sequence. For example, a regulatory region is operably linked to a coding sequence when RNA polymerase is able to transcribe the coding sequence into mRNA, which then can be translated into a protein encoded by the coding sequence.

Regulatory regions can include, without limitation, promoter sequences, enhancer sequences, response elements, protein recognition sites, inducible elements, promoter control elements, protein binding sequences, 5' and 3' untranslated regions (UTRs), transcriptional start sites, termination sequences,

polyadenylation sequences, and introns.

The nucleic acid sequence set forth in SEQ ID NO:1 is an example of a regulatory region provided herein. However, a regulatory region can have a nucleotide sequence that deviates from that set forth in SEQ ID NO:1, while
5 retaining the ability to direct expression of an operably linked nucleic acid. For example, a regulatory region having 80% or greater (*e.g.*, 81% or greater, 82% or greater, 83% or greater, 84% or greater, 85% or greater, 86% or greater, 87% or greater, 88% or greater, 89% or greater, 90% or greater, 91% or greater, 92% or greater, 93% or greater, 94% or greater, 95% or greater, 96% or greater, 97%
10 or greater, 98% or greater, or 99% or greater) sequence identity to the nucleotide sequence set forth in SEQ ID NO:1 can direct expression of an operably linked nucleic acid.

The nucleic acid sequence set forth in SEQ ID NO:2 is an example of a regulatory region provided herein. However, a regulatory region can have a
15 nucleotide sequence that deviates from that set forth in SEQ ID NO:2, while retaining the ability to direct expression of an operably linked nucleic acid. For example, a regulatory region having 80% or greater (*e.g.*, 81% or greater, 82% or greater, 83% or greater, 84% or greater, 85% or greater, 86% or greater, 87% or greater, 88% or greater, 89% or greater, 90% or greater, 91% or greater, 92%
20 or greater, 93% or greater, 94% or greater, 95% or greater, 96% or greater, 97% or greater, 98% or greater, or 99% or greater) sequence identity to the nucleotide sequence set forth in SEQ ID NO:2 can direct expression of an operably linked nucleic acid.

The term “percent sequence identity” refers to the degree of identity
25 between any given query sequence, *e.g.*, SEQ ID NO:1, and a subject sequence. A subject sequence typically has a length that is more than 80 percent, *e.g.*, more than 82, 85, 87, 89, 90, 93, 95, 97, 99, 100, 105, 110, 115, or 120 percent, of the length of the query sequence. A percent identity for any subject nucleic acid relative to a query nucleic acid can be determined as follows. A query nucleic
30 acid sequence is aligned to one or more subject nucleic acid sequences using the computer program ClustalW (version 1.83, default parameters), which allows alignments of nucleic acid sequences to be carried out across their entire length (global alignment).

ClustalW calculates the best match between a query and one or more

subject sequences, and aligns them so that identities, similarities, and differences can be determined. Gaps of one or more residues can be inserted into a query sequence, a subject sequence, or both, to maximize sequence alignments. For fast pairwise alignment of nucleic acid sequences, the following parameters are used: word size: 2; window size: 4; scoring method: percentage; number of top diagonals: 4; and gap penalty: 5. For alignment of multiple nucleic acid sequences, the following parameters are used: gap opening penalty: 10.0; gap extension penalty: 5.0; and weight transitions: yes. For fast pairwise alignment of protein sequences, the following parameters are used: word size: 1; window size: 5; scoring method: percentage; number of top diagonals: 5; and gap penalty: 3. For multiple alignment of protein sequences, the following parameters are used: weight matrix: blosum; gap opening penalty: 10.0; gap extension penalty: 0.05; hydrophilic gaps: on; hydrophilic residues: G, P, S, N, D, Q, E, R, K; and residue-specific gap penalties: on. The ClustalW output is a sequence alignment that reflects the relationship between sequences. ClustalW can be run, for example, at the Baylor College of Medicine Search Launcher site (searchlauncher.bcm.tmc.edu/multi-align/multi-align.html) and at the European Bioinformatics Institute site (ebi.ac.uk/clustalw).

To determine a percent identity between a query sequence and a subject sequence, ClustalW divides the number of identities in the best alignment by the number of residues compared (gap positions are excluded), and multiplies the result by 100. It is noted that the percent identity value can be rounded to the nearest tenth. For example, 78.11, 78.12, 78.13, and 78.14 are rounded down to 78.1, while 78.15, 78.16, 78.17, 78.18, and 78.19 are rounded up to 78.2.

A regulatory region featured herein can be made by cloning 5' flanking sequences of a fertilization-independent endosperm gene, as described in more detail below. Alternatively, a regulatory region can be made by chemical synthesis and/or PCR technology. PCR refers to a technique in which target nucleic acids are amplified. Generally, sequence information from the ends of the region of interest or beyond is employed to design oligonucleotide primers that are identical or similar in sequence to opposite strands of the template to be amplified. PCR can be used to amplify specific sequences from DNA as well as RNA, including sequences from total genomic DNA or total cellular RNA.

Primers are typically 14 to 40 nucleotides in length, but can range from 10

nucleotides to hundreds of nucleotides in length. PCR is described, for example, in *PCR Primer: A Laboratory Manual*, Ed. by Dieffenbach and Dveksler, Cold Spring Harbor Laboratory Press, 1995. Nucleic acids also can be amplified by

5 ligase chain reaction, strand displacement amplification, self-sustained sequence replication, or nucleic acid sequence-based amplification. See, for example, Lewis, *Genetic Engineering News*, 12(9):1 (1992); Guatelli *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:1874-1878 (1990); and Weiss, *Science*, 254:1292 (1991).

Various lengths of a regulatory region described herein can be made by similar
10 techniques. PCR strategies also are available by which site-specific nucleotide sequence modifications can be introduced into a template nucleic acid.

The ability of a regulatory region to direct expression of an operably linked nucleic acid can be assayed using methods known to one having ordinary skill in the art. In particular, regulatory regions of varying lengths can be
15 operably linked to a reporter nucleic acid and used to transiently or stably transform a cell, *e.g.*, a plant cell. Suitable reporter nucleic acids include β -glucuronidase (GUS), green fluorescent protein (GFP), yellow fluorescent protein (YFP), and luciferase (LUC). Expression of the gene product encoded by the reporter nucleic acid can be monitored in such transformed cells using
20 standard techniques.

When a heterologous nucleic acid is operably linked to a tissue-, organ-, or cell-preferential regulatory region, transcription occurs only or predominantly in a particular tissue, organ, and cell type, respectively. For example, a regulatory region can drive expression preferentially in a plant ovule. An ovule
25 is a structure in a flower that contains the female gametophyte and develops into a seed. A female gametophyte is also referred to in angiosperms as the embryo sac. The seed is a mature ovule, including the embryo, the endosperm, and the seed coat.

In some cases, a regulatory region can direct transcription primarily in a
30 plant ovule that has not been fertilized, such as in an un-pollinated embryo sac. In some cases, a regulatory region can direct transcription in a plant ovule that has been fertilized, such as in a plant ovule starting within 24 hours post-fertilization to at least five days (*e.g.*, six, seven, eight, nine, 10, 11, 12, 13, or 14 days) after fertilization. In some cases, a regulatory region can direct

transcription in endosperm tissue starting within 24 hours after fertilization to at least five days after fertilization.

Nucleic acid constructs

5 Nucleic acid constructs containing nucleic acids such as those described herein are also provided. A nucleic acid construct can be a vector. A vector is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements. Suitable vector backbones include, for example, those
10 routinely used in the art such as plasmids, viruses, artificial chromosomes, BACs, YACs, or PACs. The term “vector” includes cloning, transformation, and expression vectors, as well as viral vectors and integrating vectors. An expression vector is a vector that includes one or more regulatory regions.
15 Suitable expression vectors include, without limitation, plasmids and viral vectors derived from, for example, bacteriophage, baculoviruses, and retroviruses. Numerous vectors and expression systems are commercially available from such corporations as Novagen (Madison, WI), Clontech (Palo Alto, CA), Stratagene (La Jolla, CA), and Invitrogen/Life Technologies
20 (Carlsbad, CA).

 A nucleic acid construct includes a regulatory region as disclosed herein. A construct can also include a heterologous nucleic acid operably linked to the regulatory region, in which case the construct can be introduced into an organism and used to direct expression of the operably linked nucleic acid. The
25 heterologous nucleic acid can be operably linked to the regulatory region in the sense or antisense orientation. The regulatory region can be operably linked from approximately 1 to 150 nucleotides upstream of the ATG translation start codon of a heterologous nucleic acid in the sense orientation. For example, the regulatory region can be operably linked 1 nucleotide, 2 nucleotides, 3
30 nucleotides, 4 nucleotides, 5 nucleotides, 6 nucleotides, 7 nucleotides, 8 nucleotides, 9 nucleotides, 10 nucleotides, 11 nucleotides, 12 nucleotides, 13 nucleotides, 14 nucleotides, 15 nucleotides, 16 nucleotides, 17 nucleotides, 18 nucleotides, 19 nucleotides, 20 nucleotides, 25 nucleotides, 30 nucleotides, 35 nucleotides, 40 nucleotides, 45 nucleotides, 50 nucleotides, 55 nucleotides, 60

nucleotides, 65 nucleotides, 70 nucleotides, 75 nucleotides, 80 nucleotides, 85 nucleotides, 90 nucleotides, 95 nucleotides, 100 nucleotides, 110 nucleotides, 120 nucleotides, 130 nucleotides, 140 nucleotides, or 150 nucleotides upstream of the ATG translation start codon of a heterologous nucleic acid in the sense orientation. In some cases, the regulatory region can be operably linked from approximately 151 to 500 nucleotides upstream of the ATG translation start codon of a heterologous nucleic acid in the sense orientation. In some cases, the regulatory region can be operably linked from approximately 501 to 1125 nucleotides upstream of the ATG translation start codon of a heterologous nucleic acid in the sense orientation.

In some embodiments, a heterologous nucleic acid is transcribed and translated into a polypeptide. Suitable polypeptides include, without limitation, screenable and selectable markers such as green fluorescent protein, yellow fluorescent protein, luciferase, β -glucuronidase, or neomycin phosphotransferase II. Suitable polypeptides also include polypeptides that affect growth and/or hormone production. In some embodiments a heterologous nucleic acid encodes a polypeptide involved in nutrient utilization. In other embodiments, a heterologous polynucleotide encodes a non-plant protein of pharmaceutical or industrial interest. In some cases, a heterologous nucleic acid encodes a polypeptide involved in DNA methylation, such as a cytosine DNA methyltransferase. In some cases, a heterologous nucleic acid encodes a transcription factor polypeptide, such as a MADS-box transcription factor polypeptide. In some cases, a heterologous nucleic acid encodes a polypeptide involved in protein ubiquitination, such as a polypeptide having ubiquitin-protein ligase activity.

A nucleic acid construct may include a heterologous nucleic acid that is transcribed into an RNA useful for inhibiting expression of a gene. Suitable constructs from which such an RNA can be transcribed include antisense constructs. Antisense nucleic acid constructs can include a regulatory region of the invention operably linked, in antisense orientation, to a nucleic acid molecule that is heterologous to the regulatory element. Thus, for example, a transcription product can anneal to the sense coding sequence of an endogenous polypeptide. A transcription product can also be unpolyadenylated, lack a 5' cap structure, or contain an unsplicable intron. Constructs containing operably linked nucleic

acid molecules in sense orientation also can be used to inhibit the expression of a gene. Methods of co-suppression using a full-length cDNA sequence as well as a partial cDNA sequence are known in the art. See, e.g., U.S. Patent No. 5,231,020.

5 Alternatively, a heterologous nucleic acid can be transcribed into a ribozyme. See, U.S. Patent No. 6,423,885. Heterologous nucleic acid molecules can encode ribozymes designed to cleave particular mRNA transcripts, thus preventing expression of a polypeptide. Hammerhead ribozymes are useful for destroying particular mRNAs, although various ribozymes that cleave mRNA at
10 site-specific recognition sequences can be used. Hammerhead ribozymes cleave mRNAs at locations dictated by flanking regions that form complementary base pairs with the target mRNA. The sole requirement is that the target RNA contain a 5'-UG-3' nucleotide sequence. The construction and production of hammerhead ribozymes is known in the art. See, for example, U.S. Patent No.
15 5,254,678. Hammerhead ribozyme sequences can be embedded in a stable RNA such as a transfer RNA (tRNA) to increase cleavage efficiency *in vivo*. Perriman *et al.*, *Proc. Natl. Acad. Sci. USA*, 92(13):6175-6179 (1995); de Feyter and Gaudron, *Methods in Molecular Biology*, Vol. 74, Chapter 43, "Expressing Ribozymes in Plants," Edited by Turner, P.C, Humana Press Inc., Totowa, NJ.
20 RNA endoribonucleases which have been described, such as the one that occurs naturally in *Tetrahymena thermophila*, can be useful. See, for example, U.S. Patent Nos. 4,987,071 and 6,423,885.

 A nucleic acid construct also may include a heterologous nucleic acid that is transcribed into an interfering RNA. Methods for designing and preparing
25 interfering RNAs to target an endogenous gene are known to those of ordinary skill in the art. See, e.g., U.S. Patent No. 6,753,139; U.S. Patent Publication 20030175965; and U.S. Patent Publication 20030175783. A suitable interfering RNA also can be constructed as described in Brummell *et al.*, *Plant J.* 33:793-800 (2003).

30 An RNA useful for inhibiting expression of a gene can be one that can anneal to itself, *e.g.*, a double stranded RNA having a stem-loop structure. One strand of the stem portion of a double stranded RNA can comprise a sequence that is similar or identical to the sense coding sequence of the polypeptide of interest, and that is from about 10 nucleotides to about 2,500 nucleotides in

length. In some embodiments, the stem portion is similar or identical to UTR

sequences 5' of the coding sequence. In some embodiments, the stem portion is similar or identical to UTR sequences 3' of the coding sequence. The length of the sequence that is similar or identical to the sense coding sequence, the 5'

5 UTR, or the 3' UTR can be from 10 nucleotides to 50 nucleotides, from 10 nucleotides to 500 nucleotides, from 15 nucleotides to 300 nucleotides, from 20 nucleotides to 100 nucleotides, or from 25 nucleotides to 100 nucleotides. In some embodiments the length of the sequence that is similar or identical to the sense coding sequence, the 5' UTR, or the 3' UTR can be from 25 nucleotides to 10 500 nucleotides, from 25 nucleotides to 300 nucleotides, from 25 nucleotides to 1,000 nucleotides, from 100 nucleotides to 2,000 nucleotides, from 300 nucleotides to 2,500 nucleotides, from 200 nucleotides to 500 nucleotides, from 1,000 nucleotides to 2,500 nucleotides, or from 200 nucleotides to 1,000 nucleotides. The other strand of the stem portion of a double stranded RNA can 15 comprise a sequence that is similar or identical to the antisense strand of the coding sequence of the polypeptide of interest, and can have a length that is shorter, the same as, or longer than the length of the corresponding sense sequence. The loop portion of a double stranded RNA can be from 10 nucleotides to 2,500 nucleotides in length, *e.g.*, from 15 nucleotides to 100 20 nucleotides, from 20 nucleotides to 300 nucleotides, from 25 nucleotides to 400 nucleotides, or from 30 to 2,000 nucleotides in length. The loop portion of the RNA can include an intron. See, e.g., WO 97/01952; WO 98/53083; WO 99/32619; WO 98/36083; WO 99/53050; and U.S. patent publications 20040214330 and 20030180945. See also, U.S. Patents 5,034,323; 6,452,067; 25 6,777,588; 6,573,099; and 6,326,527.

If desired, a nucleic acid construct further can include a 3' untranslated region (3' UTR), which can increase stability of a transcribed sequence by providing for the addition of multiple adenylate ribonucleotides at the 3' end of the transcribed mRNA sequence. A 3' UTR can be, for example, the nopaline 30 synthase (NOS) 3' UTR. A nucleic acid construct also can contain inducible elements, intron sequences, enhancer sequences, insulator sequences, or targeting sequences other than those present in a regulatory region described herein. Regulatory regions and other nucleic acids can be incorporated into a nucleic acid construct using methods known in the art.

A nucleic acid construct may contain more than one regulatory region.

In some embodiments, each regulatory region is operably linked to a heterologous nucleic acid. For example, a nucleic acid construct may contain two regulatory regions, each operably linked to a different heterologous nucleic acid. The two regulatory regions can be the same or different, and one or both of the regulatory regions in such a construct can be a regulatory region described herein.

Transgenic Plants and Cells

Nucleic acids provided herein can be used to transform plant cells and generate transgenic plants. Thus, transgenic plants and plant cells containing the nucleic acids described herein also are provided, as are methods for making such transgenic plants and plant cells. A plant or plant cell can be transformed by having the construct integrated into its genome, *i.e.*, can be stably transformed. Stably transformed cells typically retain the introduced nucleic acid sequence with each cell division. A plant or plant cell also can be transiently transformed such that the construct is not integrated into its genome. Transiently transformed cells typically lose some or all of the introduced nucleic acid construct with each cell division, such that the introduced nucleic acid cannot be detected in daughter cells after sufficient number of cell divisions. Both transiently transformed and stably transformed transgenic plants and plant cells can be useful in the methods described herein.

Transgenic plant cells used in the methods described herein can constitute part or all of a whole plant. Such plants can be grown in a manner suitable for the species under consideration, either in a growth chamber, a greenhouse, or in a field. Transgenic plants can be bred as desired for a particular purpose, *e.g.*, to introduce a recombinant nucleic acid into other lines, to transfer a recombinant nucleic acid to other species, or for further selection of other desirable traits. Alternatively, transgenic plants can be propagated vegetatively for those species amenable to such techniques.

As used herein, a transgenic plant also refers to progeny of an initial transgenic plant. Progeny include descendants of a particular plant or plant line. Progeny of an instant plant include seeds formed on F₁, F₂, F₃, F₄, F₅, F₆, and subsequent generation plants, or seeds formed on BC₁, BC₂, BC₃, and

subsequent generation plants, or seeds formed on F₁BC₁, F₁BC₂, F₁BC₃, and subsequent generation plants. The designation F₁ refers to the progeny of a cross between two parents that are genetically distinct. The designations F₂, F₃, F₄, F₅, and F₆ refer to subsequent generations of self- or sib-pollinated progeny of an F₁ plant. Seeds produced by a transgenic plant can be grown and then selfed (or outcrossed and selfed) to obtain plants and seeds homozygous for the nucleic acid construct.

Transgenic plant cells can be grown in suspension culture, or tissue or organ culture. Solid and/or liquid tissue culture techniques can be used. When using solid medium, transgenic plant cells can be placed directly onto the medium or can be placed onto a filter film that is then placed in contact with the medium. When using liquid medium, transgenic plant cells can be placed onto a floatation device, *e.g.*, a porous membrane that contacts the liquid medium. Solid medium typically is made from liquid medium by adding agar. For example, a solid medium can be Murashige and Skoog (MS) medium containing agar and a suitable concentration of an auxin, *e.g.*, 2,4-dichlorophenoxyacetic acid (2,4-D), and a suitable concentration of a cytokinin, *e.g.*, kinetin.

Techniques for transforming a wide variety of higher plant species are known in the art. The polynucleotides and/or recombinant vectors described herein can be introduced into the genome of a plant host using any of a number of known methods, including electroporation, microinjection, and biolistic methods. Alternatively, polynucleotides or vectors can be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. Such *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well known in the art. Other gene transfer and transformation techniques include protoplast transformation through calcium or PEG, electroporation-mediated uptake of naked DNA, electroporation of plant tissues, viral vector-mediated transformation, and microprojectile bombardment (see, *e.g.*, U.S. Patents 5,538,880; 5,204,253; 5,591,616; and 6,329,571). If a cell or tissue culture is used as the recipient tissue for transformation, plants can be regenerated from transformed cultures using techniques known to those skilled in the art.

The polynucleotides and vectors described herein can be used to transform a number of monocotyledonous plants and plant cell systems, including monocots such as banana, barley, date palm, field corn, garlic, millet, oat, oil palm, onion, pineapple, popcorn, rice, rye, sorghum, sudangrass, 5 sugarcane, sweet corn, switchgrass, turf grasses, and wheat.

Thus, the methods and compositions described herein can be utilized with monocotyledonous plants such as those belonging to the orders *Alismatales*, *Arales*, *Arecales*, *Bromeliales*, *Commelinales*, *Cyclanthales*, *Cyperales*, *Eriocaulales*, *Hydrocharitales*, *Juncaceae*, *Liliales*, *Najadales*, *Orchidales*, 10 *Pandanales*, *Poales*, *Restionales*, *Triuridales*, *Typhales*, and *Zingiberales*.

The methods and compositions can be used over a broad range of plant species, including species from the monocot genera *Agrostis*, *Allium*, *Ananas*, *Andropogon*, *Asparagus*, *Avena*, *Cynodon*, *Elaeis*, *Eragrostis*, *Festuca*, *Festulolium*, *Heterocallis*, *Hordeum*, *Lemna*, *Lolium*, *Musa*, *Oryza*, *Panicum*, 15 *Pennisetum*, *Phleum*, *Phoenix*, *Poa*, *Saccharum*, *Secale*, *Sorghum*, *Triticum*, *Zea*, and *Zoysia*.

A transformed cell, callus, tissue, or plant can be identified and isolated by selecting or screening the engineered plant material for particular traits or activities, *e.g.*, those encoded by marker genes or antibiotic resistance genes. 20 Such screening and selection methodologies are well known to those having ordinary skill in the art. In addition, physical and biochemical methods can be used to identify transformants. These include Southern analysis or PCR amplification for detection of a polynucleotide; Northern blots, S1 RNase protection, primer-extension, quantitative real-time PCR, or reverse transcriptase 25 PCR (RT-PCR) amplification for detecting RNA transcripts; enzymatic assays for detecting enzyme or ribozyme activity of polypeptides and polynucleotides; and protein gel electrophoresis, Western blots, immunoprecipitation, and enzyme-linked immunoassays to detect polypeptides. Other techniques such as *in situ* hybridization, enzyme staining, and immunostaining also can be used to 30 detect the presence or expression of polypeptides and/or polynucleotides. Methods for performing all of the referenced techniques are well known.

A regulatory region disclosed herein can be used to express any of a number of heterologous nucleic acids of interest in a plant. For example, a regulatory region disclosed herein can be used to express a polypeptide or an

interfering RNA. In some cases, a regulatory region disclosed herein can be

used to express a cytosine DNA methyltransferase in female gametophyte cells

of a plant. In some cases, a regulatory region disclosed herein can be used to

express an interfering RNA that inhibits transcription of an endogenous cytosine

5 DNA methyltransferase in female gametophyte cells of a plant. Expression of

such a polypeptide or interfering RNA can affect the phenotype of a plant (*e.g.*, a

transgenic plant) when expressed in the plant, *e.g.*, at the appropriate time(s), in

the appropriate tissue(s), or at the appropriate expression levels. Thus,

transgenic plants (or plant cells) can have an altered phenotype as compared to a

10 corresponding control plant (or plant cell) that either lacks the transgene or does

not express the transgene. A corresponding control plant can be a corresponding

wild-type plant, a corresponding plant that is not transgenic but otherwise is of

the same genetic background as the transgenic plant of interest, or a

corresponding plant of the same genetic background in which expression of the

15 transgene is suppressed, inhibited, or not induced (*e.g.*, where expression is

under the control of an inducible promoter). A plant can be said "not to express"

a transgene when the plant exhibits less than 10%, *e.g.*, less than 9%, 8%, 7%,

6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.01%, or 0.001%, of the amount of the
polypeptide, mRNA encoding the polypeptide, or transcript of the transgene

20 exhibited by the plant of interest. Expression can be evaluated using methods

including, for example, quantitative real-time PCR, RT-PCR, Northern blots, S1

RNase protection, primer extensions, Western blots, protein gel electrophoresis,

immunoprecipitation, enzyme-linked immunoassays, microarray technology, and

mass spectrometry. It should be noted that if a transgene is expressed under the

25 control of a tissue-preferential or broadly expressing promoter, expression can be

evaluated in a selected tissue or in the entire plant. Similarly, if a transgene is

expressed at a particular time, *e.g.*, at a particular time during development or

upon induction, expression can be evaluated selectively during a desired time

period.

30 Use of a regulatory region provided herein to inhibit expression of an

endogenous cytosine DNA methyltransferase polypeptide in female gametophyte

cells of a plant can, after fertilization, lead to the formation of seed having an

increased average weight compared to the average weight of seed from a

corresponding control plant. In some embodiments, use of a regulatory region

provided herein to inhibit expression of an endogenous cytosine DNA

methyltransferase polypeptide in female gametophyte cells of a plant can, after fertilization, lead to the formation of seed having an increased average volume compared to the average volume of seed from a corresponding control plant. In

5 some embodiments, use of methods and compositions described herein to express a cytosine DNA methyltransferase polypeptide in female gametophyte cells of a plant can, after fertilization, lead to the formation of seed having a decreased average weight compared to the average weight of seed from a corresponding control plant. In some embodiments, use of methods and
10 compositions described herein to express a cytosine DNA methyltransferase polypeptide in female gametophyte cells of a plant can, after fertilization, lead to the formation of seed having a decreased average volume compared to the average volume of seed from a corresponding control plant.

In some embodiments, use of methods and compositions described herein
15 to express a MADS-box transcription factor polypeptide, such as a PHERES1 (PHE1) or a PHERES2 (PHE2) polypeptide, in female gametophyte cells of a plant can, after fertilization, lead to the formation of seed having an increased average weight compared to the average weight of seed from a corresponding control plant. In some embodiments, use of methods and compositions
20 described herein to express a MADS-box transcription factor polypeptide, such as a PHERES1 (PHE1) or a PHERES2 (PHE2) polypeptide, in female gametophyte cells of a plant can, after fertilization, lead to the formation of seed having an increased average volume compared to the average volume of seed from a corresponding control plant. In some embodiments, use of a regulatory
25 region provided herein to inhibit expression of an endogenous MADS-box transcription factor polypeptide, such as a PHERES1 (PHE1) or a PHERES2 (PHE2) polypeptide, in female gametophyte cells of a plant can, after fertilization, lead to the formation of seed having a decreased average weight compared to the average weight of seed from a corresponding control plant. In
30 some embodiments, use of a regulatory region provided herein to inhibit expression of an endogenous MADS-box transcription factor polypeptide, such as a PHERES1 (PHE1) or a PHERES2 (PHE2) polypeptide, in female gametophyte cells of a plant can, after fertilization, lead to the formation of seed

having a decreased average volume compared to the average volume of seed from a corresponding control plant.

In some embodiments, use of methods and compositions described herein to express a polypeptide annotated as having ubiquitin-protein ligase activity, such as a MEIDOS (MEO) polypeptide, in female gametophyte cells of a plant can, after fertilization, lead to the formation of seed having an increased average weight compared to the average weight of seed from a corresponding control plant. In some embodiments, use of methods and compositions described herein to express a polypeptide annotated as having ubiquitin-protein ligase activity, such as a MEIDOS (MEO) polypeptide, in female gametophyte cells of a plant can, after fertilization, lead to the formation of seed having an increased average volume compared to the average volume of seed from a corresponding control plant. In some embodiments, use of a regulatory region provided herein to inhibit expression of an endogenous polypeptide annotated as having ubiquitin-protein ligase activity, such as a MEIDOS (MEO) polypeptide, in female gametophyte cells of a plant can, after fertilization, lead to the formation of seed having a decreased average weight compared to the average weight of seed from a corresponding control plant. In some embodiments, use of a regulatory region provided herein to inhibit transcription of an endogenous polypeptide annotated as having ubiquitin-protein ligase activity, such as a MEIDOS (MEO) polypeptide, in female gametophyte cells of a plant can, after fertilization, lead to the formation of seed having a decreased average volume compared to the average volume of seed from a corresponding control plant.

Seeds of transgenic plants describe herein can be conditioned and bagged in packaging material by means known in the art to form an article of manufacture. Packaging material such as paper and cloth are well known in the art. Such a bag of seed preferably has a package label accompanying the bag, *e.g.*, a tag or label secured to the packaging material, a label printed on the packaging material, or a label inserted within the bag. The package label may indicate the seed contained therein incorporates transgenes that provide increased seed weight.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1 – Isolation of a 5'-flanking region of a fertilization-independent endosperm gene

Rice gene expression profiles (Lan *et al.*, *Plant Mol Biol.* 54(4):471-87 (2004)) were analyzed to identify genes that were highly expressed in pistil five days after pollination and that were not expressed in un-pollinated pistil. A fertilization-independent endosperm (fie2) gene was identified as such a gene. The expression level of fie2 in pistil five days after pollination was about 28-fold higher than the expression level of fie2 in unpollinated pistil.

The sequence of the fie2 expressed sequence tag (EST) used by Lan *et al.* to construct a cDNA microarray was retrieved from the website of the National Center for Gene Research, Chinese Academy of Sciences (ncgr.ac.cn/EST.html). The fie2 EST sequence was then compared to sequences in the National Center for Biotechnology Information (NCBI) database using the Basic Local Alignment Search Tool (BLAST), and a corresponding complementary DNA (cDNA) sequence was identified. The cDNA sequence was then compared to cDNA clones in the database of full-length cDNA clones from japonica rice (Knowledge-based Oryza Molecular biological Encyclopedia: KOME; cdna01.dna.affrc.go.jp/cDNA/) to identify a predicted full-length cDNA sequence. The predicted full-length cDNA sequence was used to perform a BLAST search and retrieve the corresponding genomic DNA sequence. By aligning the genomic DNA sequence with the coding sequence, the ATG start codon and the 5'-flanking region were identified. Approximately three kilobases of the 5'-flanking region were isolated. This sequence, designated pOsFie2-2 and predicted to have the sequence set forth in SEQ ID NO:1, was cloned into an expression vector such that it was operably linked to a Histone-Yellow Fluorescent Protein (YFP) expression cassette that had previously been tested using a 35S promoter.

Example 2 – Analysis of pOsFie2-2 activity

Rice of the Kitaake cultivar was transformed with an expression vector containing a Histone-YFP coding sequence under the transcriptional control of pOsFie2-2. Ten putative transformants were selected and screened for the presence of the transgene using the polymerase chain reaction (PCR) with gene

specific primers. The transgene was present in all ten transformants. The transgene was isolated from transformed plants and sequenced. The *in planta* OsFie2-2 sequence is set forth in SEQ ID NO:2.

Whole ovules were collected from four of the transgenic plants.

5 Approximately ten pre-fertilization ovules were dissected from the pistil of each plant, flash frozen in liquid nitrogen, and pooled for RNA extraction. Post-fertilization ovules were collected separately and processed in a similar manner. RNA samples were extracted from pre- and post-fertilization ovules and analyzed using reverse transcription PCR (RT-PCR). Plants in which expression
10 of the Histone-YFP fusion protein was detected by RT-PCR were analyzed further using confocal microscopy. At least four ovules were dissected from each plant that was positive for Histone-YFP expression according to the RT-PCR assay. Isolated post-fertilization (one to two days after pollination) and pre-fertilization ovules were analyzed for Histone-YFP expression using
15 confocal microscopy with different light channels. Ovules were examined using a YFP channel, a chlorophyll channel, and a bright field.

Microscopy analysis indicated that pOsFie2-2 was active as a promoter in early stages of seed development. YFP expression was observed in five out of five transformed plants analyzed. YFP expression was observed in seeds as
20 early as 24 hours after fertilization. These results suggest that pOsFie2-2 is active as a promoter in the earliest stages of endosperm development, from fertilization to about three days after fertilization. Results from other experiments indicated that pOsFie2-2 was not active as a promoter in endosperm at 14 days after fertilization.

25 In addition to analyzing post-fertilization ovules, pre-fertilization ovules arising from plants transformed with the expression vector containing the Histone-YFP coding sequence under the transcriptional control of pOsFie2-2 were analyzed. Microscopy analysis carried out as described above indicated that YFP expression also occurred in the embryo sac prior to fertilization.

30 Taken together, these results indicate that pOsFie2-2 is active in ovules prior to fertilization and remains active until at least three days after fertilization.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined

5 by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid comprising a regulatory region having 80 percent or greater sequence identity to SEQ ID NO:1 or SEQ ID NO:2, wherein said regulatory region directs transcription, in a plant ovule within 24 hours post-fertilization, of an operably linked heterologous polynucleotide.
2. The nucleic acid of claim 1, wherein said regulatory region has at least 85 percent or greater sequence identity to SEQ ID NO:1 or SEQ ID NO:2.
3. The nucleic acid of claim 1, wherein said regulatory region has at least 90 percent or greater sequence identity to SEQ ID NO:1 or SEQ ID NO:2.
4. The nucleic acid of claim 1, wherein said regulatory region comprises an intron.
5. An isolated nucleic acid comprising a regulatory region having 80 percent or greater sequence identity to SEQ ID NO:1 or SEQ ID NO:2, wherein said regulatory region directs transcription in an unfertilized plant ovule of an operably linked heterologous polynucleotide.
6. The nucleic acid of claim 5, wherein said regulatory region has at least 85 percent or greater sequence identity to SEQ ID NO:1 or SEQ ID NO:2.
7. The nucleic acid of claim 5, wherein said regulatory region has at least 90 percent or greater sequence identity to SEQ ID NO:1 or SEQ ID NO:2.
8. The nucleic acid of claim 5, wherein said regulatory region comprises an intron.
9. A nucleic acid construct comprising the nucleic acid of claim 1 or claim 5 operably linked to a heterologous polynucleotide.

10. The nucleic acid construct of claim 9, wherein said heterologous polynucleotide comprises a nucleic acid sequence encoding a polypeptide.
11. The nucleic acid construct of claim 9, wherein said heterologous polynucleotide is in an antisense orientation relative to said regulatory region.
12. The nucleic acid construct of claim 9, wherein said heterologous polynucleotide is transcribed into an antisense RNA capable of inhibiting expression of a DNA methyltransferase.
13. The nucleic acid construct of claim 9, wherein said heterologous polynucleotide is transcribed into an interfering RNA.
14. The nucleic acid construct of claim 9, wherein said heterologous polynucleotide is transcribed into an interfering RNA against a DNA methyltransferase.
15. A transgenic plant or plant cell transformed with the nucleic acid of claim 1 or claim 5 operably linked to a heterologous polynucleotide.
16. The transgenic plant or plant cell of claim 15, wherein said heterologous polynucleotide comprises a nucleic acid sequence encoding a polypeptide.
17. The transgenic plant or plant cell of claim 15, wherein said heterologous polynucleotide is in an antisense orientation relative to said regulatory region.
18. The transgenic plant or plant cell of claim 15, wherein said heterologous polynucleotide is transcribed into an interfering RNA.
19. A method of producing a transgenic plant, said method comprising (a) introducing into a plant cell an isolated polynucleotide comprising the nucleic acid of claim 1 or claim 5 operably linked to a heterologous polynucleotide, and (b) growing a plant from

said plant cell.

20. The method of claim 19, wherein said heterologous polynucleotide comprises a nucleic acid sequence encoding a polypeptide.

21. The method of claim 19, wherein said heterologous polynucleotide is in an antisense orientation relative to said regulatory region.

22. The method of claim 19, wherein said heterologous polynucleotide is transcribed into an interfering RNA.

DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT
(PCT Article 17(2)(a), Rules 13ter.1(c) and (d) and 39)

Applicant's or agent's file reference 18207-011WO1	IMPORTANT DECLARATION	Date of mailing (<i>day/month/year</i>) 08 OCTOBER 2007 (08.10.2007)
International application No. PCT/US2007/060235	International filing date (<i>day/month/year</i>) 08 JANUARY 2007 (08.01.2007)	(Earliest) Priority date (<i>day/month/year</i>) 09 JANUARY 2006 (09.01.2006)
International Patent Classification (IPC) or both national classification and IPC C12N 15/29(2006.01)i		
Applicant CERES, INC. et al		

This International Searching Authority hereby declares, according to Article 17(2)(a), that **no international search report will be established** on the international application for the reasons indicated below.

1. ☐ The subject matter of the international application relates to:
 - a. ☐ scientific theories.
 - b. ☐ mathematical theories.
 - c. ☐ plant varieties.
 - d. ☐ animal varieties.
 - e. ☐ essentially biological processes for the production of plants and animals, other than microbiological processes and the products of such processes.
 - f. ☐ schemes, rules or methods of doing business.
 - g. ☐ schemes, rules or methods of performing purely mental acts.
 - h. ☐ schemes, rules or methods of playing games.
 - i. ☐ methods for treatment of the human body by surgery or therapy.
 - j. ☐ methods for treatment of the animal body by surgery or therapy.
 - k. ☐ diagnostic methods practised on the human or animal body.
 - l. ☐ mere presentation of information.
 - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art.
2. ☒ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:

☐ the description
 ☒ the claims
 ☐ the drawings
3. ☐ A meaningful search could not be carried out without the sequence listing; the applicant did not, within the prescribed time limit:

☐ furnish a sequence listing on paper complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in a form and manner acceptable to it.
☐ furnish a sequence listing in electronic form complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in a form and manner acceptable to it.
☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13ter.1(a) or (b)
4. ☐ A meaningful search could not be carried out without the tables related to the sequence listings; the applicant did not, within the prescribed time limit, furnish such tables in electronic form complying with the technical requirements provided for in Annex C-bis of the Administrative Instructions, and such tables were not available to the International Searching Authority in a form and manner acceptable to it.
5. Further comments:

Name and mailing address of ISA/KR  Korean Intellectual Property Office 920 Dunsan-dong, Seo-gu, Daejeon 302-701, Republic of Korea Facsimile No. 82-42-472-7140	Authorized officer JEONG Eui Jun  Telephone No. 82-42-481-5549
---	--